6 Statistiek en beslissingen

Voorkennis Spreiding en steekproeven

Bladzijde 47

1 **a** $\bar{x} \approx 19{,}9$ en $\sigma \approx 1{,}6$ geeft $\bar{x} - \sigma \approx 19{,}9 - 1{,}6 = 18{,}3$ en $\bar{x} + \sigma \approx 19{,}9 + 1{,}6 = 21{,}5$.
b Er zijn 6 + 9 + 5 = 20 leerlingen met een aantal ademhalingen per minuut tussen 18,3 en 21,5.
c De totale frequentie is 1 + 1 + 3 + 6 + 9 + 5 + 3 + 2 = 30.
Dus van $\frac{20}{30} \times 100\% \approx 66{,}7\%$.

2 **a** Merk A
Voer in lijst 1 = {103, 104, 105, 106, 107, 108, 109, 110}
en lijst 2 = {3, 15, 28, 36, 48, 39, 12, 4}.
1-Var Stats (TI) of 1VAR (Casio) geeft $\bar{x} = 106{,}6$ en $\sigma \approx 1{,}5$.
Merk B
Voer in lijst 2 = {1, 3, 11, 42, 53, 16, 3, 1}.
1-Var Stats (TI) of 1VAR (Casio) geeft $\bar{x} = 106{,}6$ en $\sigma \approx 1{,}1$.
Dus bij de steekproef van merk A is het gemiddelde 106,6 gram en de standaardafwijking 1,5 gram en bij de steekproef van merk B is het gemiddelde 106,6 gram en de standaardafwijking 1,1 gram.
b Bij merk B is de vulmachine beter afgesteld, want er is minder spreiding.

6

3 Voer in lijst 1 = {9,1; 10,3; 11,8; 14,1; 17,4; 21,2; 24,4; 24,3; 21,7; 17,5; 13,1; 10,0} en lijst 2 = {−3,1; −0,6; 4,4; 11,5; 16,1; 20,2; 23,4; 22,8; 18,1; 11,9; 5,7; 0,3}.
1-Var Stats (TI) of 1VAR (Casio) geeft met lijst 1 en frequentie 1 dat $\bar{x} \approx 16{,}2$ en $\sigma \approx 5{,}4$ en met lijst 2 en frequentie 1 dat $\bar{x} \approx 10{,}9$ en $\sigma \approx 9{,}0$.
Dus bij Barcelona is het gemiddelde 16,2 °C en de standaardafwijking 5,4 °C.
Bij Boekarest is het gemiddelde 10,9 °C en de standaardafwijking 9,0 °C.

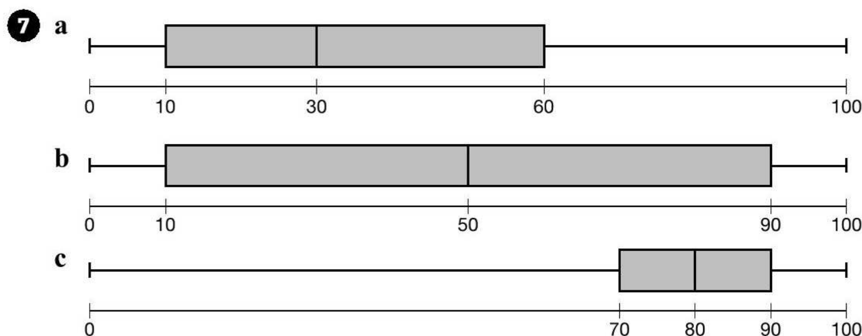
4 **a** Bij beiden is de gemiddelde bezorgtijd 40 minuten.
Bij Ben liggen de meeste bezorgtijden bij 39, 40 en 41 (dus maximaal één minuut naast het gemiddelde) en maar 2 van de 24 tijden twee minuten naast het gemiddelde.
Bij Annet is de spreiding veel groter, er liggen 16 van de 24 tijden twee minuten naast het gemiddelde.
Dus bij Annet is de spreiding het grootst.
b Schatting bezorgtijd Ben: $\sigma = 0{,}8$ minuten.
Schatting bezorgtijd Annet: $\sigma = 1{,}7$ minuten.

Bladzijde 48

5 **a** Voer in list 1 = {15, 16, 17, 18, 19, 20, 21} en lijst 2 = {8, 3, 1, 7, 2, 6, 12}.
1-Var Stats (TI) of 1VAR (Casio) geeft $\bar{x} \approx 18{,}49$ en $\sigma \approx 2{,}32$.
Meer dan 1,6 van het gemiddelde geeft waarnemingsgetallen kleiner dan 18,5 − 1,6 = 16,9 en groter dan 18,5 + 1,6 = 20,1.
Dus bij de waarnemingsgetallen 15, 16 en 21 en hierbij horen 8 + 3 + 12 = 23 waarnemingen van in totaal 8 + 3 + 1 + 7 + 2 + 6 + 12 = 39 waarnemingen.
Dit geeft $\frac{23}{39} \times 100\% \approx 59{,}0\%$.
b $\bar{x} - \sigma \approx 18{,}49 - 2{,}32 = 16{,}17$ en $\bar{x} + \sigma \approx 18{,}49 + 2{,}32 = 20{,}81$
Hiertussen liggen 1 + 7 + 2 + 6 = 16 waarnemingen van de 39.
Dit geeft $\frac{16}{39} \times 100\% \approx 41{,}0\%$.

6 **a** Boxplot A en B zijn beide symmetrisch rond de mediaan 50, dus hierbij zullen II en IV horen.
Bij A is de spreiding minder groot dan bij B,
dus bij A horen gemiddelde = 50 en standaardafwijking = 22 (IV).
Bij B horen gemiddelde = 50 en standaardafwijking = 38 (II).
Bij D is 75% van de waarnemingen groter dan 50 terwijl bij C 75% groter is dan 40. Bij D verwacht je dus eerder een hoger gemiddelde dan andersom. Bij D is de kwartielafstand kleiner dan bij C, je verwacht dus ook dat de standaardafwijking bij D kleiner is dan bij C.
Dus bij C horen gemiddelde = 60 en standaardafwijking = 27 (I).
Bij D horen gemiddelde = 65 en standaardafwijking = 23 (III).

7



Bladzijde 49

8 Stel er zijn in de steekproef x leerlingen met een bijbaantje.

In het geval $\hat{p} = p$ geldt $\frac{x}{65} = \frac{380}{1415}$

$$1415 \cdot x = 380 \cdot 65$$

$$x = \frac{380 \cdot 65}{1415} \approx 17$$

In de steekproef zitten 17 leerlingen met een bijbaantje.

9 **a** Er worden 712 huishoudens ondervraagd, dus $\frac{435}{712} \approx 0{,}61$ is een steekproefproportie.

b Er geldt $\frac{86}{435} \times 100\% \approx 19{,}8\%$, maar hierbij is uitgegaan van de 435 huishoudens die aangeven regelmatig op internet te kopen.

Zou je uitgaan van de 712 ondervraagde huishoudens, dan krijg je $\frac{86}{712} \approx 0{,}12$.

Dus 12% is een betere schatting.

c 6% van 435 is ongeveer 26.
Maar de 435 huishoudens geven aan regelmatig aankopen via internet te doen.
Daarom zal bij deze 435 huishoudens het percentage hoger zijn dan 6%.
Daarom is keuzemogelijkheid III (meer dan 30) de juiste.

6.1 Soorten verdelingen

Bladzijde 50

1 **a** Er waren 7 wedstrijden met 2 doelpunten.
b Er zijn in totaal $8 + 4 + 7 + 2 + 3 + 6 + 2 + 1 + 1 = 34$ wedstrijden gespeeld.
c Er zijn in totaal $0 \cdot 8 + 1 \cdot 4 + 2 \cdot 7 + 3 \cdot 2 + 4 \cdot 3 + 5 \cdot 6 + 6 \cdot 6 + 7 \cdot 1 + 8 \cdot 1 = 93$ doelpunten gescoord.
d Een wedstrijd met eindstand 2-2 is een wedstrijd met 4 doelpunten. Dus er zijn maximaal 3 wedstrijden in 2-2 geëindigd.
e De modus is 0 en het gemiddelde is zeker groter dan 0, want rechts van de staaf bij 0 staan ook staven die aantallen doelpunten groter dan nul aangeven.
Dus het gemiddelde is groter dan de modus.

Bladzijde 51

2 **a** Het gemiddelde in figuur 6.8 zal kleiner zijn dan het gemiddelde in figuur 6.6.
b De mediaan is in beide figuren hetzelfde.
c De standaardafwijking in figuur 6.8 zal groter zijn dan de standaardafwijking in figuur 6.6.

Bladzijde 52

3 **a** Er is uitgegaan van de klassenbreedte 2 cm.
b Jongens zijn over het algemeen langer dan meisjes. In de linker top zitten vooral meisjes en in de rechter top vooral jongens.
c De modale klasse is de klasse 182-<184.
d Het gemiddelde is ongeveer 176 cm. De lengten wijken hier gemiddeld ongeveer 8 cm van af. Een schatting van de standaardafwijking is dus 8 cm.
e Als boxplot I bij het histogram zou passen dan zou 25% van de waarnemingen tussen 160 en 164 liggen, dat is in figuur 6.9 duidelijk niet het geval.
Er zitten rechts van 176 cm een stuk meer waarnemingsgetallen dan links van 176 cm. De mediaan zal dus groter zijn dan 176 cm. Boxplot III past dus niet bij het histogram.
Boxplot II past het beste bij figuur 6.9, want ongeveer een kwart van de waarnemingen zit tussen 179 cm en 183 cm.

Bladzijde 53

4 **a** Nee, de modale klasse is 12 000 tot 18 000. De mediaan ligt rechts van deze klasse.
b Nee, als iedereen een gelijk inkomen heeft is dat het enige inkomen dat voorkomt en bestaat de verdeling dus uit één staaf.
c In figuur 6.11 is te zien dat bijna 25% van de jaarinkomens lager is dan 12 000 euro. Hierdoor vallen boxplots II en III duidelijk af. Boxplot I past dus het beste.

Bladzijde 54

5 **a** symmetrisch
b rechts-scheef
c rechts-scheef
d symmetrisch
e rechts-scheef

6 Bij I hoort E.
Bij II hoort B.
Bij III hoort A.
Bij IV hoort C.
Bij V hoort D.

Bladzijde 55

7 **a** Veel geboorten vinden plaats in het ziekenhuis, dus nuljarigen maken veel kosten.
b De staaf bij de klasse 55-59 is de eerste die boven 1360 euro uitkomt. Dus vanaf de leeftijdsgroep 55-59 jaar zijn de kosten meer dan gemiddeld.
c De kosten zijn per persoon. Irma gaat ervan uit dat elke klasse evenveel personen bevat, maar dat blijkt niet uit figuur 6.14.
d De klassenbreedtes zijn niet gelijk. De klasse van de nuljarigen is 1 jaar breed, de meeste andere klassen zijn 5 jaar breed.
e 45-49 kosten $= 1000 \cdot 1\,300\,000 = 1{,}3$ milard euro
75-79 kosten $= 4000 \cdot 450\,000 = 1{,}8$ miljard euro
De groep 75-79 heeft dus de hoogste totale kosten.
f Zie vraag e, de totale kosten zijn in de klasse 75-79 hoger dan in de klasse 45-49. De modale klasse ligt dus nogal aan de rechterkant van het histogram. Er is dus sprake van een links-scheve verdeling. Mark heeft dus gelijk.

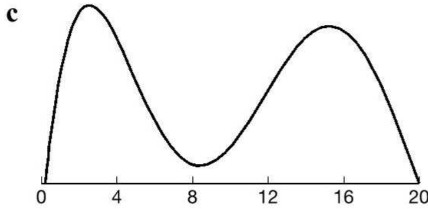
Bladzijde 56

8 De modus zal kleiner zijn dan de mediaan, want de modus zit vrijwel helemaal links en dan heb je nog niet de helft van de waarnemingsgetallen.

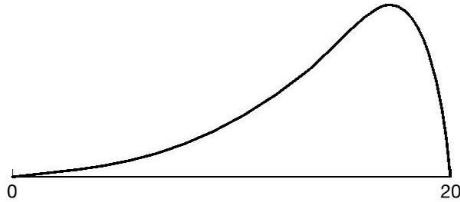
Bladzijde 57

9 **a** Bij de top van de verdelingskromme hoort de grootste frequentie, dus de cumulatieve verdelingskromme neemt daar het meest toe ofwel daar is het steilste stuk van de verdelingskromme.
b De cumulatieve verdelingskromme is overal even steil, de verdelingskromme is dus overal even hoog, dus er is sprake van een uniforme verdeling.

10 **a** Aflezen bij 50% geeft de mediaan 12.
b Er zijn twee steile gedeelten in de cumulatieve verdelingskromme, dus er is sprake van een tweetoppige verdeling.
c



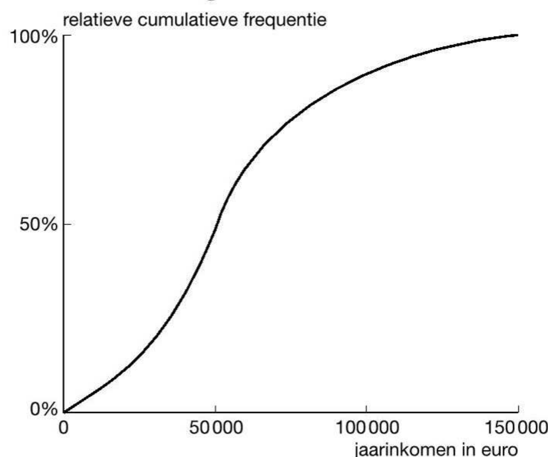
11 **a** Weinig waarnemingsgetallen in het begin en veel op het eind.



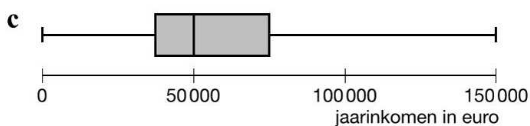
b Het is een links-scheve verdeling.
c Het gemiddelde ligt links van de mediaan.

6

12 **a** De mediaan ligt rechts van de modus.
b Het steilste stuk zit ongeveer bij 25 000 euro.
De mediaan is ongeveer 50 000 euro.



c



6.2 De normale verdeling

Bladzijde 59

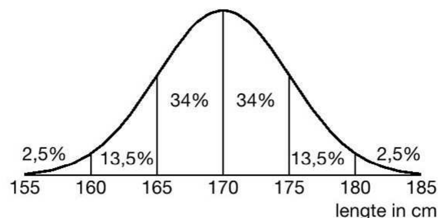
13 **a** Er is uitgegaan van een klassenbreedte van 5 cm.
b De groep bestaat uit 15 + 80 + 235 + 370 + 210 + 80 + 10 = 1000 personen.
c Dat is $\frac{680}{1000} \times 100\% = 68\%$.
d Dat is $\frac{950}{1000} \times 100\% = 95\%$.

14 **a** De klassenbreedte is 1 cm.
b De frequentie van de klasse 182 -< 183 is ongeveer 375.
c Nee, bij figuur 6.24 is de groep veel groter.
In figuur 6.24 is het aantal mannen met een lengte tussen 180 en 185 cm al ongeveer $5 \cdot 350 = 1750$.

Bladzijde 61

15 Normale verdeling bij a, c, e en f.

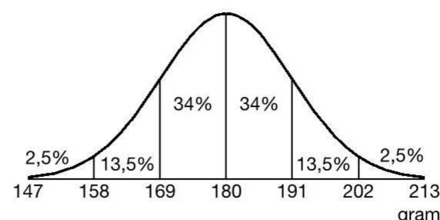
16 **a**



b 34% + 34% + 13,5% = 81,5%
c 2,5%
d 13,5% + 2,5% = 16%
e 13,5% + 34% = 47,5%

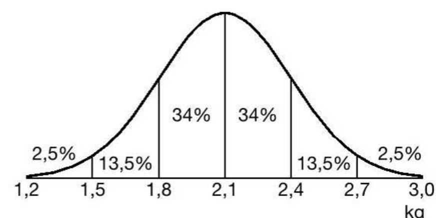
Bladzijde 62

17 **a** 13,5% + 34% + 34% = 81,5%
Dus $0{,}815 \cdot 5000 = 4075$ appels.
b 84%, dus $0{,}84 \cdot 5000 = 4200$ appels.
c 125 is $\frac{125}{5000} \times 100\% = 2{,}5\%$.
Dus zwaarder dan 202 gram.



18 16% weegt minder dan 76 gram, dus 76 ligt één standaardafwijking van het gemiddelde af.
De standaardafwijking is 80 − 76 = 4 gram.

19 **a** 2,5%
b 13,5% + 68% = 81,5%, dus $0{,}815 \cdot 200 = 163$ konijnen.
c 2,5% + 13,5% = 16%, dus $0{,}16 \cdot 200 = 32$ konijnen.
d $\frac{5}{200} \times 100\% = 2{,}5\%$, dus deze hebben een gewicht
van meer dan 2,7 kg.



20 Om in een van de buitenste vakjes terecht te komen moet een knikker óf steeds naar links óf steeds naar rechts vallen.
Er is maar één route naar een buitenste vakje. Naar de middelste vakjes zijn juist heel veel routes.

21 **a** Met stijgende leeftijd neemt iemands reactietijd toe. Bij de 18-jarigen hoort kromme A (kleinste gemiddelde), bij de 60-jarigen hoort kromme C (grootste gemiddelde).
b Bij kromme C hoort de grootste standaardafwijking, dus bij 60-jarigen is het genoemde percentage het grootst.

Bladzijde 63

22 **a** D, B, A, C
b Kromme A: $\mu = 65$ en $\sigma = \frac{70-60}{6} \approx 1{,}7$.

Kromme B: $\mu = 66{,}5$ en $\sigma = \frac{70-63}{6} \approx 1{,}2$.

Kromme C: $\mu = 67{,}5$ en $\sigma = \frac{74-61}{6} \approx 2{,}2$.

Kromme D: $\mu = 70$ en $\sigma = \frac{71{,}5-68{,}5}{6} = 0{,}5$.

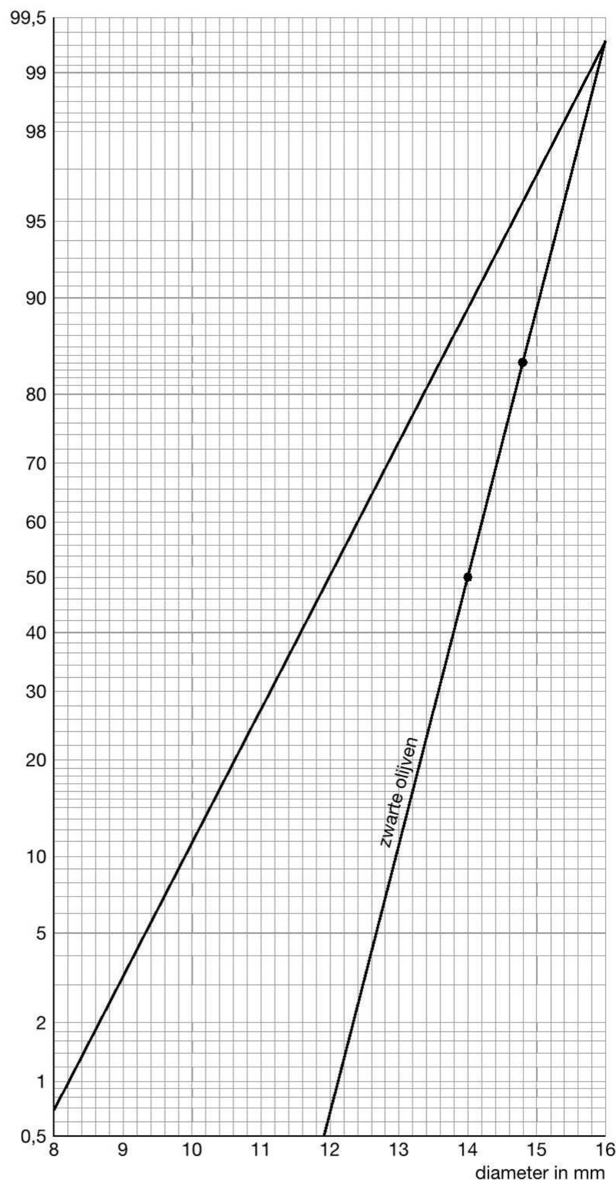
23 **a** Lees af bij 50%. Je krijgt $\mu \approx 7{,}8$.
b Bij $\mu + \sigma$ hoort 50% + 34% = 84%.
Aflezen geeft $\mu + \sigma \approx 8{,}9$.
c $\sigma \approx 8{,}9 - 7{,}8 = 1{,}1$

Bladzijde 66

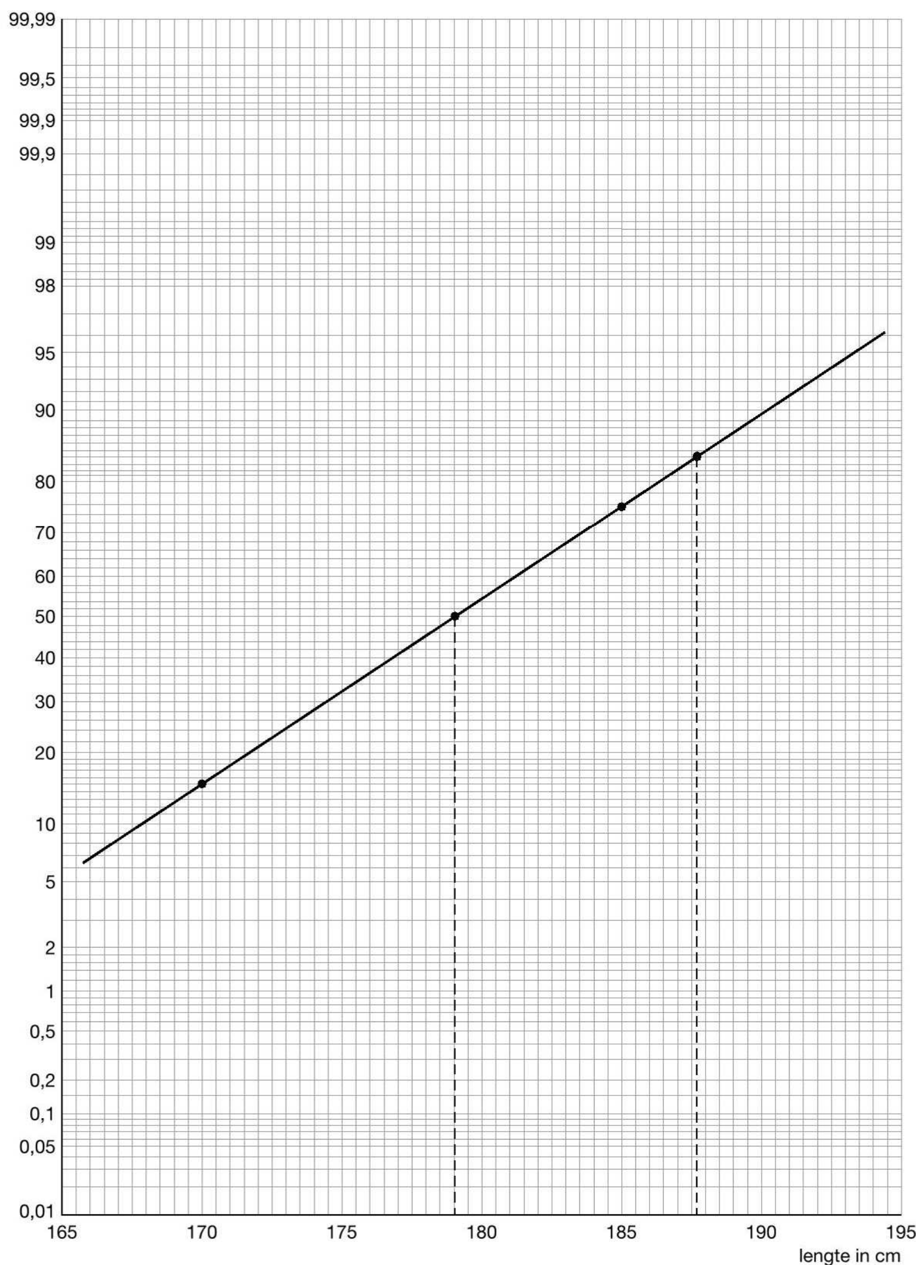
24 **a** Bij 10 mm hoort 11%.
Bij 13 mm hoort 74%.
Dus 74% − 11% = 63%.

b Lees af: bij 80% hoort ongeveer 13,3 mm.
Dus de grootste 20% van de olijven heeft een diameter van meer dan 13,3 mm.

c Bij de groene olijven is $\mu = 12$ mm.
Lees af bij 84% dat $\mu + \sigma = 13{,}6$ mm, dus $\sigma = 1{,}6$ mm.
Dus bij de zwarte olijven is $\mu = 14$ mm en $\sigma = 0{,}8$ mm.
Teken dus de lijn door (14, 50) en (14,8; 84).



25 Bij 170 cm hoort 15%.
Bij 185 cm hoort 75%.
Lees af bij 50% dat $\mu \approx 179$ cm.
Lees af bij 84% dat $\mu + \sigma \approx 188$, dus $\sigma \approx 188 - 179 = 9$ cm.



26 **a** Evenwijdig betekent dezelfde standaardafwijking.
b Dat de bladlengte van soort *A* een kleinere standaardafwijking heeft dan die van soort *C*.
c Dat zowel bij soort *C* als bij soort *D* 80% van de bladeren korter is dan 45 mm.
d De lijnen bij *B* en *D* moeten elkaar dan snijden op een hoogte van 50.

6.3 Betrouwbaarheidsintervallen

Bladzijde 68

27 **a** De populatieproportie is het gedeelte van de populatie met een bepaald kenmerk, en dat gedeelte is hier 12% ofwel 0,12.
b *
Vind je bijvoorbeeld 14, dan is de steekproefproportie 0,14.
c *
d *
e Bij een grotere steekproeflengte liggen de steekproefproporties tussen nauwere grenzen.
f Tussen 0,09 en 0,15.
g Dat zijn de grenzen 0,10 en 0,14.

Bladzijde 70

28 **a** $n = 41$ en $\hat{p} = p = 0{,}18$ geeft $\sigma = \sqrt{\dfrac{0{,}18 \cdot 0{,}82}{41}} = 0{,}06$.

b Er geldt $\mu = p = 0{,}18$ en $\sigma = 0{,}06$, dus $\mu - 2\sigma = 0{,}18 - 0{,}12 = 0{,}06$ en $\mu + 2\sigma = 0{,}18 + 0{,}12 = 0{,}30$.
Volgens een van de vuistregels van de normale verdeling heeft 95% van de steekproeven een steekproefproportie tussen 0,06 en 0,30.

c 64 is 32% van 200, dus je hebt met de vuistregel bij de normale verdeling te maken die over 68% gaat.
De 64 steekproeven hebben een steekproefproportie die minder is dan $0{,}18 - 0{,}06 = 0{,}12$ of meer dan $0{,}18 + 0{,}06 = 0{,}24$.

Bladzijde 71

29 **a** $n = 125$ en $\hat{p} = p = 0{,}28$ geeft $\sigma = \sqrt{\dfrac{0{,}28 \cdot 0{,}72}{125}} = 0{,}040159...$
De standaardafwijking is ongeveer 4,02%.

b $\dfrac{25}{125} = 0{,}20$ en $\dfrac{45}{125} = 0{,}36$
Er geldt $\mu = p = 0{,}28$ en $\sigma = 0{,}04$, dus $\mu - 2\sigma = 0{,}28 - 0{,}08 = 0{,}20$ en $\mu + 2\sigma = 0{,}28 + 0{,}08 = 0{,}36$.
Volgens een van de vuistregels van de normale verdeling heeft 95% van de steekproeven een steekproefproportie tussen 0,20 en 0,36.
Dat zijn $0{,}95 \cdot 760 = 722$ steekproeven.

c $\dfrac{40}{125} = 0{,}32$
$\mu + \sigma = 0{,}28 + 0{,}04 = 0{,}32$
Dus bij 16% van de steekproeven.
Dat zijn dus $0{,}16 \cdot 760 \approx 122$ steekproeven.

d $\dfrac{25}{125} = 0{,}20$ en $\mu - 2\sigma = 0{,}28 - 0{,}08 = 0{,}20$
Dus bij 2,5% van de steekproeven.
Dat zijn er $0{,}025 \cdot 760 = 19$.

30 **a** $n = 117$ en $\hat{p} = p = 0{,}120$ geeft $\sigma = \sqrt{\dfrac{0{,}120 \cdot 0{,}880}{117}} = 0{,}0300...$
Dus het percentage van 3,0% is juist.

b $\dfrac{18}{117} = 0{,}1538...$
Er geldt $\mu = 0{,}120$ en $\sigma = 0{,}030$, dus $\mu + \sigma = 0{,}120 + 0{,}030 = 0{,}150$.
Dus bij ongeveer 16% van de steekproeven.
Dat zijn $0{,}16 \cdot 365 \approx 58$ steekproeven.

c $\dfrac{7}{117} = 0{,}0598...$
$\mu - 2\sigma = 0{,}120 - 0{,}060 = 0{,}060$
Dus bij 2,5% van de steekproeven.
Dat zijn $0{,}025 \cdot 365 \approx 9$ dagen.

d $\dfrac{2}{117} = 0{,}0170...$
Omdat $\mu - 3\sigma = 0{,}120 - 0{,}090 = 0{,}030$ en $0{,}017 < 0{,}030$ zal het vrijwel nooit voorkomen dat in een steekproef hoogstens 2 pakken zitten met te weinig inhoud.
Dus het kan niet kloppen wat de bedrijfsleider beweert.

31 Bij $n = 80$ en $\hat{p} = 0{,}75$ is $\sigma = \sqrt{\dfrac{0{,}75 \cdot 0{,}25}{80}} \approx 0{,}05$.
Omdat $0{,}75 - 0{,}05 = 0{,}70$ en $0{,}75 + 0{,}05 = 0{,}80$ kan het ook best zijn dat minder dan 70% of meer dan 80% dezelfde favoriete sport heeft.

Bladzijde 74

32 **a** $\hat{p} - \sigma = 0{,}63 - 0{,}013... \approx 0{,}616$
$\hat{p} + \sigma = 0{,}63 + 0{,}013... \approx 0{,}644$
Het 68%-betrouwbaarheidsinterval is [0,616; 0,644].

b $\hat{p} = \frac{378}{600} = 0{,}63$ en $\sigma = \sqrt{\frac{0{,}63 \cdot 0{,}37}{600}} = 0{,}0197... \approx 0{,}020$

c $\hat{p} - 2\sigma = 0{,}63 - 2 \cdot 0{,}0197... \approx 0{,}591$
$\hat{p} + 2\sigma = 0{,}63 + 2 \cdot 0{,}0197... \approx 0{,}669$
Het 95%-betrouwbaarheidsinterval is [0,591; 0,669].

d Bij de steekproeflengte 1200 hoort een kleiner 95%-betrouwbaarheidsinterval dan bij steekproeflengte 600.
Dus bij steekproeflengte 1200 hoort een grotere nauwkeurigheid dan bij steekproeflengte 600.

Bladzijde 75

33 **a** $\hat{p} = \frac{235}{840} = 0{,}2797...$ en $\sigma = \sqrt{\frac{0{,}2797... \cdot 0{,}7202...}{840}} = 0{,}0154...$
$\hat{p} - 2\sigma = 0{,}2797... - 2 \cdot 0{,}0154... \approx 0{,}249$
$\hat{p} + 2\sigma = 0{,}2797... + 2 \cdot 0{,}0154... \approx 0{,}311$
Het 95%-betrouwbaarheidsinterval is [0,249; 0,311].

b $\hat{p} = \frac{146}{470} = 0{,}3106...$ en $\sigma = \sqrt{\frac{0{,}3106... \cdot 0{,}6893...}{470}} = 0{,}0213...$
$\hat{p} - 2\sigma = 0{,}3106... - 2 \cdot 0{,}0213... \approx 0{,}268$
$\hat{p} + 2\sigma = 0{,}3106... + 2 \cdot 0{,}0213... \approx 0{,}353$
Het 95%-betrouwbaarheidsinterval is [0,268; 0,353].

c Van de vrouwen rookten er 235 − 146 = 89.

$\hat{p} = \frac{89}{370} = 0{,}2405...$ en $\sigma = \sqrt{\frac{0{,}2405... \cdot 0{,}7594...}{370}} = 0{,}0222...$
$\hat{p} - 2\sigma = 0{,}2405... - 2 \cdot 0{,}0222... \approx 0{,}196$
$\hat{p} + 2\sigma = 0{,}2405... + 2 \cdot 0{,}0222... \approx 0{,}285$
Het 95%-betrouwbaarheidsinterval is [0,196; 0,285].

34 **a** $\hat{p} = \frac{47}{575} = 0{,}0817...$ en $\sigma = \sqrt{\frac{0{,}0817... \cdot 0{,}9182...}{575}} = 0{,}0114...$
$\hat{p} - 2\sigma = 0{,}0817... - 2 \cdot 0{,}0114... \approx 0{,}059$
$\hat{p} + 2\sigma = 0{,}0817... + 2 \cdot 0{,}0114... \approx 0{,}105$
Het 95%-betrouwbaarheidsinterval is [0,059; 0,105].

b $4\sigma = 0{,}296 - 0{,}224 = 0{,}072$
Dus de standaardafwijking is $\sigma = \frac{0{,}072}{4} = 0{,}018$.

35 De steekproefproportie van de mensen die zeggen op A te gaan stemmen is $\hat{p} = \frac{487}{935} = 0{,}5208...$

De bijbehorende standaardafwijking is $\sigma = \sqrt{\frac{0{,}5208... \cdot 0{,}4791...}{935}} = 0{,}0163...$

$\hat{p} - 2\sigma = 0{,}5208... - 2 \cdot 0{,}0163... \approx 0{,}488$
Om de verkiezingen te winnen zal kandidaat A meer dan 50% van de stemmen moeten krijgen ($p > 0{,}5$).
Omdat $0{,}488 < 0{,}5$ kun je niet met een betrouwbaarheid van 95% zeggen dat kandidaat A de verkiezingen zal gaan winnen.

36 $\hat{p} = 0{,}60$ geeft $\sigma = \sqrt{\frac{0{,}60 \cdot 0{,}40}{n}} = \sqrt{\frac{0{,}24}{n}}$

Omdat $\sigma = 2\% = 0{,}02$ krijg je $\sqrt{\frac{0{,}24}{n}} = 0{,}02$.

6

Bladzijde 76

37 $4\sigma = 0{,}46 - 0{,}34 = 0{,}12$, dus $\sigma = 0{,}03$.

$\hat{p} = \dfrac{0{,}34 + 0{,}46}{2} = 0{,}40$

$\sigma = \sqrt{\dfrac{0{,}40 \cdot 0{,}60}{n}} = \sqrt{\dfrac{0{,}24}{n}}$

Je krijgt de vergelijking $\sqrt{\dfrac{0{,}24}{n}} = 0{,}03$.

Voer in $y_1 = \sqrt{\dfrac{0{,}24}{x}}$ en $y_2 = 0{,}03$.

Intersect geeft $x \approx 267$, dus de steekproefomvang is 267.

38 **a** $p = 0{,}35$ geeft $\sigma = \sqrt{\dfrac{0{,}35 \cdot 0{,}65}{n}} = \sqrt{\dfrac{0{,}2275}{n}}$

Je krijgt de vergelijking $\sqrt{\dfrac{0{,}2275}{n}} = 0{,}025$.

Voer in $y_1 = \sqrt{\dfrac{0{,}2275}{x}}$ en $y_2 = 0{,}025$.

Intersect geeft $x = 364$.
De omvang van de steekproef moet minstens 364 zijn.

b $n = 364$ en $p = 0{,}38$ geeft $\sigma = \sqrt{\dfrac{0{,}38 \cdot 0{,}62}{364}} = 0{,}0254...$
Nu is $\sigma > 0{,}025$, dus dan was een steekproefomvang van 364 niet voldoende geweest.

39 **a** Als het een 95%-betrouwbaarheidsinterval betreft is $4\sigma = 0{,}61 - 0{,}49 = 0{,}12$, dus $\sigma = 0{,}03$.

$\hat{p} = 0{,}55$ geeft $\sigma = \sqrt{\dfrac{0{,}55 \cdot 0{,}45}{n}} = \sqrt{\dfrac{0{,}2475}{n}}$

Je krijgt de vergelijking $\sqrt{\dfrac{0{,}2475}{n}} = 0{,}03$.

Voer in $y_1 = \sqrt{\dfrac{0{,}2475}{x}}$ en $y_2 = 0{,}03$.
Intersect geeft $x = 275$, dus dan zijn 275 D66-stemmers ondervraagd.

b Als het een 68%-betrouwbaarheidsinterval betreft is $\sigma = \dfrac{0{,}12}{2} = 0{,}06$.

Je krijgt de vergelijking $\sqrt{\dfrac{0{,}2475}{n}} = 0{,}06$.

Voer in $y_1 = \sqrt{\dfrac{0{,}2475}{x}}$ en $y_2 = 0{,}06$.
Intersect geeft $x \approx 69$, dus dan zijn 69 D66-stemmers ondervraagd.

40 Stel de steekproefproportie van de ondervraagden die het examen niet te moeilijk vond is $\hat{p}$.

Je krijgt de vergelijking $\sqrt{\dfrac{\hat{p}(1 - \hat{p})}{850}} = 0{,}012$.

Voer in $y_1 = \sqrt{\dfrac{x(1 - x)}{850}}$ en $y_2 = 0{,}012$.
Neem bijvoorbeeld Xmin = 0, Xmax = 0,3, Ymin = 0 en Ymax = 0,024.
Intersect geeft $x = 0{,}1427...$, dus $\hat{p} = 0{,}1427...$
Het aantal ondervraagden dat het examen niet te moeilijk vond is $0{,}1427... \cdot 850 \approx 121$.

6.4 Groepen en kenmerken

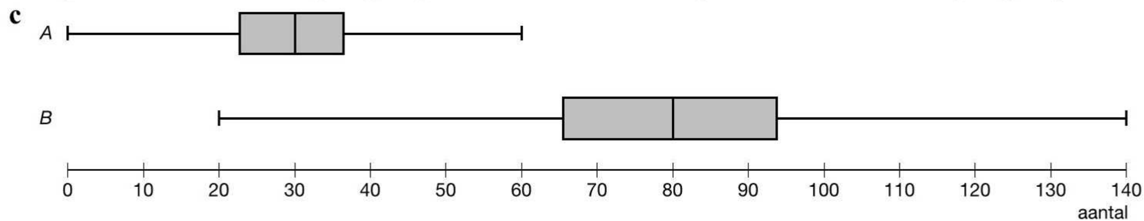
Bladzijde 78

41 **a** Het zijn rechts-scheve verdelingen.
b Verdeling I heeft het grootste gemiddelde, want bij verdeling I zijn meer waarnemingen groter dan 100 en minder waarnemingen kleiner dan 100 dan bij verdeling II.
Verdeling II heeft de grootste standaardafwijking, want daar zijn de waarnemingen meer gespreid dan bij verdeling I.

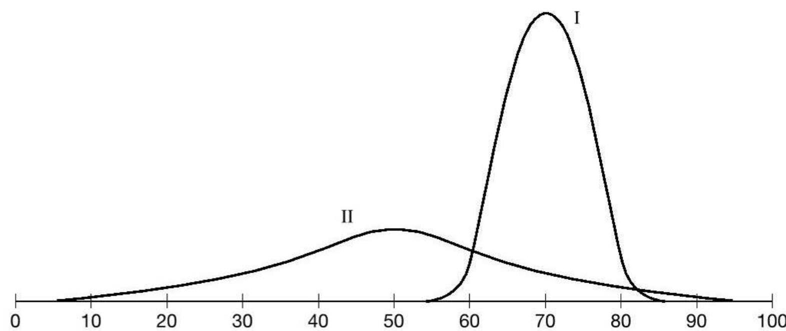
Bladzijde 79

42 Boxplot II hoort bij verdeling A, want verdeling A heeft een langere staart.

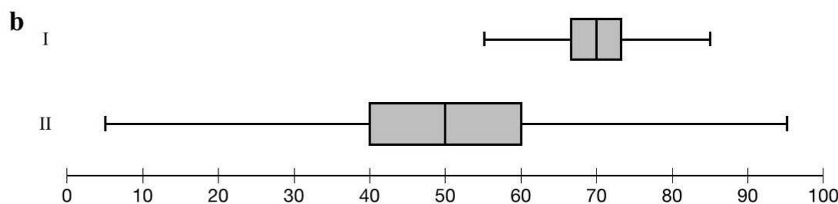
43 **a** De oppervlakte onder kromme B is groter dan de oppervlakte onder kromme A, dus bij B horen meer waarnemingsgetallen dan bij B.
b Bij A is de standaardafwijking ongeveer 30 − 20 = 10 en bij B is de standaardafwijking ongeveer 80 − 60 =20.
c



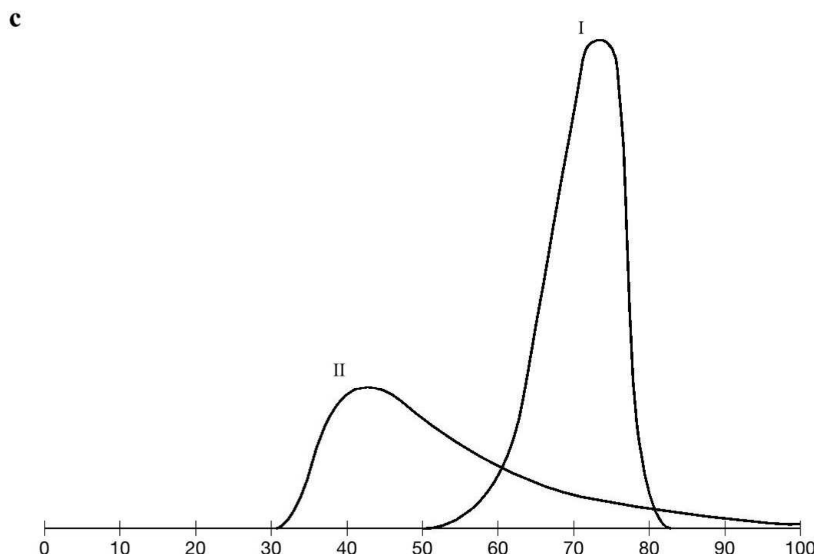
44 **a** I $\mu - 3\sigma = 70 - 3 \cdot 5 = 55$ en $\mu + 3\sigma = 70 + 3 \cdot 5 = 85$
II $\mu - 3\sigma = 50 - 3 \cdot 15 = 5$ en $\mu + 3\sigma = 50 + 3 \cdot 15 = 95$
De oppervlakte onder I is twee keer zo groot als de oppervlakte onder II.



b



c



Bladzijde 80

45 **a** In de figuur staat de procentuele verdeling, dus je weet niets van de absolute bedragen.
Ik ben het niet eens met Hendrik-Jan.

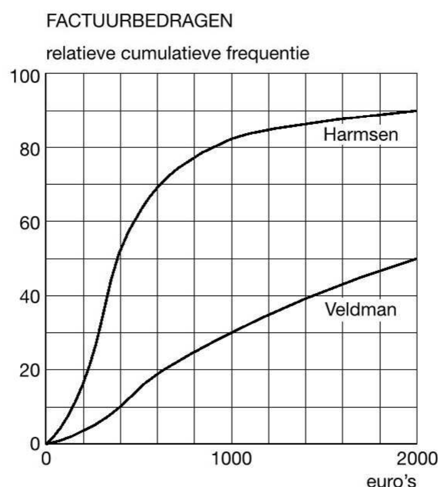
b Stel Harmsen heeft in totaal 100 opdrachten uitgevoerd.
Dan had hij 35 opdrachten uit categorie A die hem minder dan $35 \cdot 250 = 8750$ euro opleverden.
En dan had hij 20 opdrachten uit categorie C die hem in totaal minimaal $20 \cdot 500 = 10\,000$ euro opleverden.
Dus hij had meer omzet in categorie C dan in categorie A.

c Van de opdrachten van Veldman had 50% een factuurbedrag tot 2000 euro, de mediaan is dus 2000.
Dus P hoort bij Veldman en Q bij Harmsen.

d Deze kromme hoort bij hovenier Harmsen, immers bij de relatieve cumulatieve frequentie 50 hoort een factuurbedrag van iets minder dan 400 euro en dat hoort bij Harmsen.

e Bij Veldman hoort de volgende tabel.

rel. cum. freq.	5	16	30	50
factuurbedrag in euro's	250	500	1000	2000



Bladzijde 81

46 Nee, het kan ook juist andersom zijn. Ook kan het zijn dat beide een gevolg zijn van een derde (niet genoemde) oorzaak. De onderzoekers kunnen dus niet zomaar zeggen dat het één een gevolg is van het ander.

Bladzijde 82

47 **a** het warme weer

b de leeftijd

c de toegenomen welvaart

d de lichaamslengte

48 **a** Geen causaal verband. Aan voorwaarde 2 (de tijd) is niet voldaan.

b Wel een causaal verband.

c Geen causaal verband. Aan voorwaarde 3 is niet voldaan. Bij een grote brand zijn meer brandweerlieden aanwezig, maar die veroorzaken de grote brand niet.

d Wel een causaal verband.

e Geen causaal verband. Aan voorwaarde 3 is niet voldaan. In een straat met veel BMWs zijn veel rijke gezinnen en die verbruiken wellicht meer energie.

49 Als de dokter bijna alleen mannelijke patiënten heeft, is het percentage ongeveer 22%, bij vrijwel alleen vrouwelijke patiënten 18%. Voor een mix van mannen en vrouwen zal het percentage dus tussen de 18% en 22% liggen. Dus conclusie 3 kun je trekken en conclusie 1 niet.
Omdat geen absolute aantallen bekend zijn, kun je de conclusies 2 en 4 niet trekken.

Bladzijde 84

50 Alleen bij puntenwolk a zou de conclusie van de onderzoeker wel juist zijn, want hierbij zijn zowel in de vakken recht onder en recht boven de gekleurde rechthoek geen punten en ook niet in de vakken links en rechts van de gekleurde rechthoek.

51 1 Ja, want de lijn die bij 50 hoort is vrijwel een rechte lijn.
2 Ja, want de lijnen lopen steeds verder uit elkaar.
3 Nee, na 36 weken is de mediaan van de gewichten ongeveer 2600 gram. Dus dan weegt de helft van de jongens minder dan 2600 gram en de helft meer dan 2600 gram.
4 Nee, in week 28 weegt 10% van de jongens 1500 gram of meer en in week 36 weegt 10% van de jongens ongeveer 3000 gram of meer. Dat is iets anders dan dat 90% van de jongens is gegroeid van 1500 naar 3000 gram.
5 Ja, deze conclusie is juist. Zie conclusie 4.

Bladzijde 85

52 Dat kan hij niet concluderen. Mensen die vlees eten hebben wellicht een andere levensstijl dan vegetariërs. De enige conclusie uit het artikel kan zijn dat mensen met een hoge bloeddruk wellicht baat hebben bij een vegetarisch dieet.

53 Dat kan ze niet concluderen. Depressieve mensen besteden wellicht minder aandacht aan de persoonlijke verzorging, dus misschien is het juist wel omgekeerd, dus dat depressiviteit een slecht gebit veroorzaakt.

54 Dat zijn 123inkt.nl, Navulwinkel en Q-Nomic. Bij deze drie leveranciers is de prijs per mL laag, terwijl de tevredenheid goed tot zeer goed is.

Bladzijde 86

55 **a** Bij $G = 0$ is oppervlakte A gelijk aan nul en oppervlakte B niet gelijk is aan nul.
Omdat de formule bij $A = 0$ als uitkomst $G = 0$ moet geven komen alleen de formules I en III in aanmerking (de uitkomst van een breuk is nul als de teller nul is).
Omdat bij $B = 0$ moet horen $G = 1$ is formule III juist. Immers $B = 0$ geeft dan $G = \frac{A}{A+0} = 1$.

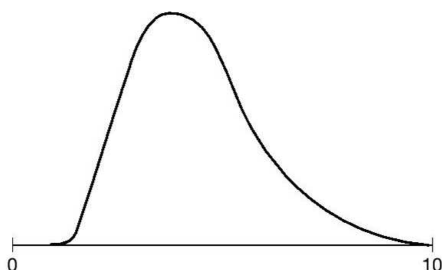
b 1 Nee, de Gini-coëfficiënt van het besteedbaar inkomen is laag.
2 Ja, de Gini-coëfficiënt van het besteedbaar inkomen schommelt tussen 0,273 en 0,281.
3 Ja, dat klopt. Zo geldt voor 2012 bijvoorbeeld dat $\frac{0{,}540}{0{,}274} \approx 1{,}97$.
4 Nee, de tabel gaat over inkomens en niet over vermogens.
5 In 2010 was deze herverdeling $\frac{0{,}249}{0{,}530} \times 100\% \approx 47{,}0\%$ en in 2012 was het $\frac{0{,}266}{0{,}540} \times 100\% \approx 49{,}3\%$.
Dus de conclusie is juist.

Diagnostische toets

Bladzijde 88

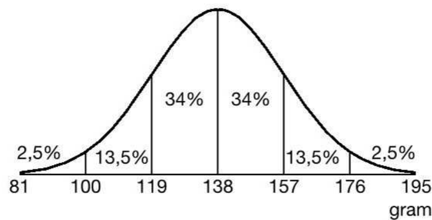
1 **a** Een links-scheve verdeling en een tweetoppige verdeling.
b Er zijn veel leerlingen die de woordjes heel slecht hebben geleerd en er zijn veel leerlingen die de woordjes heel goed hebben geleerd.
c gemiddelde 4,6
mediaan 2
standaardafwijking 4,0
d Bijvoorbeeld cijfer 100 intikken in plaats van 10.

2 **a** Bij relatieve cumulatieve frequentie 50 hoort het waarnemingsgetal 2,8. Dus de mediaan is ongeveer 2,8.
b Veel waarnemingen aan het begin en weinig op het eind, dus een rechts-scheve verdeling.
c



d Nee, de mediaan is ongeveer 2,8 (zie vraag a) en dit is bij de boxplot niet het geval.

3 **a**



Meer dan 157 gram is 13,5% + 2,5% = 16%.

b 2,5%, dat zijn dus $0{,}025 \cdot 120 = 3$ bananen.

c Het gemiddelde is $\dfrac{90 + 222}{2} = 156$ gram.

$6\sigma = 222 - 90 = 132$ gram, dus de standaardafwijking is $\sigma = \dfrac{132}{6} = 22$ gram.

Bladzijde 89

4 *A* heeft het kleinste gemiddelde, dus III hoort bij *A*.
Van *B* en *C* heeft *C* de kleinste standaardafwijking (de lijn die bij *C* hoort is steiler dan de lijn die bij *B* hoort), dus I hoort bij *C*.
Dus bij *A* hoort III, bij *B* hoort II en bij *C* hoort I.

5 **a** $\hat{p} = \dfrac{16}{80} = 0{,}2$ geeft $\sigma = \sqrt{\dfrac{\hat{p}(1 - \hat{p})}{n}} = \sqrt{\dfrac{0{,}2 \cdot 0{,}8}{80}} = 0{,}0447...$

In één decimaal nauwkeurig is $\sigma = 4{,}5\%$.

b $\mu + 2\sigma = 0{,}2 + 2 \cdot 0{,}045 = 0{,}29$

Dus 30% past niet binnen het 95%-betrouwbaarheidsinterval.

c $\sigma = \sqrt{\dfrac{0{,}17 \cdot 0{,}83}{80}} \approx 0{,}042$

$\mu + \sigma = 0{,}17 + 0{,}042 = 0{,}212$

Je verwacht 16% van de scholen met een percentage van 21% of hoger.
16% van 25 is 4
Dus 4 scholen met een percentage van 21% of hoger is niet uitzonderlijk.

d Breedte 0,16 betekent $4\sigma = 0{,}16$, dus $\sigma = 0{,}04$.

Stel *n* leerlingen per school ondervragen.

Er geldt dan $\sigma = \sqrt{\dfrac{0{,}17 \cdot 0{,}83}{n}} = \sqrt{\dfrac{0{,}1411}{n}}$.

Je krijgt de vergelijking $\sqrt{\dfrac{0{,}1411}{n}} = 0{,}04$.

Voer in $y_1 = \sqrt{\dfrac{0{,}1411}{x}}$ en $y_2 = 0{,}04$.

Intersect geeft $x \approx 88{,}2$, dus minstens 89 leerlingen per school ondervragen.

6 I Deze conclusie kun je trekken. In 4HA is dat ongeveer 65% en in 4HB is dat 60%.
II Die conclusie kun je niet trekken, omdat je geen absolute aantallen weet.
III Deze conclusie kun je niet trekken. Om deze conclusie te kunnen trekken, moet je weten hoe de verdeling binnen de cijferklassen is.

7 Bij c is geen sprake van een causaal verband, want agressieve jongeren spelen wellicht vaker gewelddadige computergames.
Bij d is geen sprake van een causaal verband, want er zijn veel meer oorzaken waardoor de levensverwachting is toegenomen.

